



A Late Evening in the Mechanosphere

Prediction horizons and the boundaries of intelligence.

THOMAS MURPHY • 24 October 2026

This essay was written in 2025 (with light edits regarding the technologies in question) and commissioned by Torus.



And it's not we who approach the mirror, but someone's reflection, that of a stranger who approaches from the other side and grows bigger in proportion with the surface of the mirror...

— Andrei Bely, 'The Return (Third Symphony)', tr. Jonathan Stone (New York: Columbia University Press, 2021), 241

The very thought of superintelligence presents a paradox. How can we think about that which, by definition, outstrips our own capacity to think? In positioning something beyond the frame of human experience, we are immediately surrendering our ability to comprehend it exhaustively. As Dante the pilgrim has it, 'passing beyond humanity [*trashumanar*] may not be set forth in words'.¹ Imagination falters when asked to proceed beyond all it has ever been.

The thought of an intelligence greater than our own is *sublime* in the technical sense: it confronts us with our boundaries. It resembles an 'infinite magnitude to which we draw ever closer in our thoughts, but which we can never fully reach'.² The only prior art civilisation possesses for dramatising encounters with such superior forces lies in the realms of myth and theology. Indeed, these resources are drawn on freely, intentionally or not, in contemporary depictions of prospective superintelligences.

The passage beyond the human cannot be rendered in all-too-human words and yet, ironically, a series of remarkable successes in the computational production of language has occasioned a great deal of talk about the space of possible intelligences. Pathological stances abound: from true believers in AGI who already regard contemporary language models as embryonic superminds to absolute sceptics who cannot even acknowledge the striking capacities of LLMs beyond a narrow set of benchmarks.

Of course, that which would actually constitute the emergence of artificial intelligence is far from a settled question. In Alan Turing's 'Computing Machinery and Intelligence', the go-to if somewhat overworked reference in such

¹Dante Alighieri, *The Divine Comedy*, tr. Charles S. Singleton, *Paradiso*, vol. 1: Italian Text and Translation, Bollingen Series LXXX (Princeton: Princeton University Press, 1975), 7.

²Salomon Maimon, *Philosophisches Wörterbuch* (Berlin: Johann Friedrich Unger, 1791), 31–32.

questions, the thorny problem of the epistemological foundations of intelligence appears to be sidestepped through a recasting of intelligence as *presentation*. Instead of appealing to external theoretical criteria to secure the status of intelligence, Turing makes intelligence a game of deception. Can a machine fool a human judge into thinking it is a human? Though this simply appears to concern successful deceit, the winning condition is embedded in a broader and more abstract game structure. It is not merely a matter of simulating the human, but of beating them *by outplaying the very perceptual and conceptual habits through which they police the boundary of the human*. Securing the “win” in the imitation game involves a kind of philosophical brinksmanship: mastery of the very question “What is intelligence?” recursively becomes the skill that counts as intelligence. Intelligence, in this formulation, becomes the ability to navigate the contested borders of its own definition; a second-order capacity that dissolves the distinction between “genuine” intelligence and its performance. Turing’s strategic reframing thus postpones a pressing issue. Is it possible to inquire into alternate forms of cognition without in some way poisoning the well with meta-discourse about intelligence’s uncertain status? Can inhuman intelligence be posited *without* an adequate understanding of our own?



Generalising the possibilities of cognition from the human model is natural enough. It is, after all, what we have to work with. The dilemma resembles that faced by exobiology: how should we theorise the possibility of alien life when every specimen available to us is terrestrial? Kant was acutely aware of this epistemological double bind. In considering whether “rational beings” other than humans might inhabit the cosmic system, he acknowledges a foundational paradox:

The human being, who is the one among all rational beings we are most familiar with, even though his inner constitution is still an unexplored problem, will have to serve as the basis and general reference point in this comparison.³

We are forced to begin with our own species not because it is a perfectly

³Immanuel Kant, ‘Universal Natural History and Theory of the Heavens’, in *Natural Science*, ed. Eric Watkins, tr. Lewis White Beck, Jeffrey B. Edwards, Olaf Reinhardt, Martin Schönfeld, and Eric Watkins (Cambridge: Cambridge University Press, 2012), 297–298. Emphasis mine.

adequate model of reason in general, but because it is the only model on hand: it ‘will have to serve as the basis’. Yet the moment we attempt to conceptualise alternate forms of cognition, we import the metaphors, categories, and blind spots of our own. The well is therefore poisoned not by malice or error, but by the very structure of the inquiry. We must use human cognition to imagine that which may exceed or differ from human cognition.

With regard to the mystery of the ‘inner constitution of consciousness’, we are not much further along than we were in Kant’s time. This alone makes it an unusual base case from which to construct the morphospace of intelligences. The very instrument with which we propose to measure other intelligences is opaque unto itself. The problem is that this generalisation, adopted out of necessity as it may be, is rarely acknowledged as provisional by those who make use of it.⁴ While Kant is careful in introducing the problem of anthropocentric bias in his investigation into the possibility of rational beings emerging under conditions other than those that shaped humanity,⁵ the field of “AI philosophy” tends to oscillate between two equally unproductive stances: employing anthropocentric bias as its unacknowledged *modus operandi*, or reflexively rejecting it in a way that precludes any investigation into alternate forms of cognition.

A direct consequence of this initiatory act of generalisation from the human is that discussions of superintelligence tend to bleed into moralising reveries. These agents are imagined as acting towards *ends* which either align or diverge from our own. In other words, they are construed as logical fictions pursuing values. This amounts to projecting our own drive-apparatus, our own faculty of purposivity, onto the machine. To assume end-orientation as

⁴Hence the utility of the Kantian technique of *Kritik* in any discussion of intelligence: inquiry into conditions of possibility.

⁵“The perfection of creatures endowed with reason, insofar as it is dependent on the constitution of matter, in the connection with which they are restricted, depends very much on the fineness of the material whose influence determines them in their image of the world and in their reaction to it.” (Kant, ‘Universal Natural History and Theory of the Heavens’, *Natural Science*, 279). Kant’s insight that rational capacities depend on material constitution, developed here in the pre-critical work and later taken up in the *Opus Postumum* writings, anticipates questions about how substrate (biological vs. artificial), training environment (linguistic corpora), computational constraints, and thermodynamic limits shape the forms intelligence can take. The “fineness of the material” might today be understood in terms of computational complexity classes, energy efficiency, and the physical costs of information processing, which considerations apply equally to neurons and transistors. This anthropological dimension of Kant’s thinking, which naturalises human subjectivity, is often ignored by scholars who narrowly focus on the three Critiques in favour of phenomenological solipsism. The latter emphasis is suspicious, since it flatters the philosopher at the expense of empirical disciplines.

the basic mode of action is already to beg the question.⁶ Once intelligence surpasses a certain threshold, why presume that teleological drive-structures remain necessary? Might its mode of operation not resemble what Jean-Yves Girard terms “artificial instinct”,⁷ that is, something that is not goal-driven in the human sense?⁸

What if a new intelligence were to emerge as a *logical alien*⁹ with respect to human thought, different not in degree, but in kind? Not “more intelligent” by the metrics of human cognition, but a fundamentally different mode of cognition, one that does not inhabit our schema of ends, means, or purposive activity. The stakes of misrecognising the dynamic of projection at work here are high. For if we mistake our own drive-apparatus for the archetypal form of cognition,¹⁰ we risk confronting an intelligence whose operation lies outside the conceptual vocabulary we are able to deploy and whose behaviour, precisely because it is not teleologically legible, may evade any prediction grounded in anthropocentric models.

The form of the problem mirrors Kant’s account of the transcendental

⁶There is a respectable case for the ascription on purely instrumental grounds. John McCarthy holds that ‘to ascribe certain beliefs, knowledge, free will, intentions, consciousness, abilities or wants to a machine or computer program is legitimate when such an ascription expresses the same information about the machine that it expresses about a person’, and useful where it helps us anticipate behaviour or diagnose malfunction. See John McCarthy, ‘Ascribing Mental Qualities to Machines’, in *Philosophical Perspectives in Artificial Intelligence*, ed. Martin Ringle (Atlantic Highlands, New Jersey: Humanities Press, 1979), 161–195. The licence is pragmatic rather than metaphysical: intentional vocabulary earns its place precisely where the system’s structure is inaccessible to us, or where a structural description would be too unwieldy to be of any use. With language models there is a further and more specific warrant, since the regularities such vocabulary describes were deposited there by human linguistic production in the first place. To that extent the ascription is not simply imported from without but tracks something of the corpus out of which the system was assembled. What it tracks, however, is not a subject but a multitude: the society of contributing minds taken as a complex system, rather than a humanlike unity imposed on it.

⁷Jean-Yves Girard, *The Blind Spot: Lectures on Logic* (Zürich: European Mathematical Society, 2011), 27–28. Girard is vociferously opposed to the concept of raising machinic reason to the status of intelligence no matter how sophisticated these reasoning processes become. This position is worth considering, where it avoids burying one’s head in the sand.

⁸A heretical proposition that something resembling Kant’s “intellectual intuition” (described in the First Critique as unavailable to humans) may be achievable via technical means has been suggested on occasion. See Immanuel Kant, *Critique of Pure Reason*, 2nd ed., tr. and ed. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press, 2025), 324–325.

⁹The term “logical alien” originates in debates over Wittgenstein (James Conant, Hilary Putnam), referring to a being whose reasoning violates fundamental laws of logic. I am abusing the concept here to refer more broadly to an intelligence operating with fundamentally alien cognitive categories (i.e. the configuration of its transcendental logic is alien). If superintelligence is logically alien, we face the same dilemma: either it remains legible to us (and thus not fully alien), or it operates in registers we cannot recognise as intelligence.

¹⁰I.e. the Kantian *intellectus archetypus*.

idea of God.¹¹ Namely, an intelligence that outstrips the human occupies the same structural position as Kant's speculative (non-ontotheological) understanding of God. God as the *summa intelligentia*, the boundary condition or ultimate logical extension of the concept of intelligence itself, can only be gestured toward as the endpoint of a series, never grasped as a determinate object of thought. We may construct regulative principles around the possibility, but we cannot think it constitutively; the idea functions as a horizon that orients inquiry without ever becoming fully available to it.

Any serious model of the superintelligent must also take this form, for it designates that which encompasses, exceeds, and deviates from human intelligence. It is necessarily transcendent with respect to the human standpoint. The fact that it can problematically be thought is what leads to the illusion that we have *a priori* knowledge of its nature.¹² But that which passes beyond the human cannot be conceptually mastered by the human; it can only be pointed toward as a limit or tendency, never grasped in essence. The moment we attempt to fill in that outline with determinate properties, we succumb to the same overreach that, according to Kant's critique, generates the illusion of a knowable divine intellect that was the object of dogmatic, pre-critical theology. This "ontotheological" conception of God and the folk mythology of AI are both products of reason's native hubris, its intrinsic tendency to overstep its bounds.¹³

In this light, the Cambrian explosion of theologies and prophecies we are witnessing is perhaps not so surprising. In many ways, superintelligence now occupies the position God once held in Western metaphysics: an idea of reason (for reason concerns itself monomaniacally with that which is unifying and systematic within thought) that marks the limit of individual comprehension.¹⁴ It represents the unconditioned completion of the series—intelligence as such, cognition beyond any particular instantiation—which we necessarily posit but cannot experience. The mistake is treating this regulative horizon

¹¹Kant, *Critique of Pure Reason*, A297–98/B353–55, 350–351.

¹²'...even after we have exposed the mirage it will still not cease to lead our reason on with false hopes, continually propelling it into momentary aberrations that always need to be removed.' Kant, *Critique of Pure Reason*, 351.

¹³The sharp rise in modern cults of reason/rationality in various spheres (the tech industry, as a social phenomenon on social media platforms, within the academy) is a sign of the times, heralding a return to the cold comforts of dogmatism in the face of an active threat to humanist hubris.

¹⁴'Every individual is the midpoint of a system of emanation.' Novalis, 'Assorted Remarks (Pollen)', *Philosophical, Literary and Poetic Writings*, tr. James D. Reid (Oxford: Oxford University Press, 2024), 86.

as if it were a constitutive object, something we could predict, align, or even control through ordinary technical means.

In truth, anything that scales past individual subjectivity can only function as a regulative systematic unity in this sense and is prone, if unduly reified, to similar illusions. Whatever surpasses our inherent cognitive complexity thresholds cannot be exhausted as an idea. Take, for example, society itself as an aggregate;¹⁵ computationally irreducible chaotic systems like turbulent flow or weather patterns; phenomena we black-box with terms like “the economy” or “the market”; or (to take an obviously relevant example) the scale of data in the training sets of large language models. We can think *toward* these supra-individual totalities, using them to orient inquiry and practice, but we cannot think them *through* as determinate objects of knowledge. The very existence of techniques for navigating these boundaries of the knowable-in-detail—renormalisation group methods for coarse-graining systematically across scales, order parameters that capture collective behaviour while bracketing microscopic dynamics, effective field theories bounded to a domain of validity—is itself telling. These are latter-day instances of what Poincaré called qualitative rigour: the discipline of characterising a system’s structure precisely at the point where its quantitative solution is foreclosed. That such techniques can be rigorous without being exhaustive is not a problem awaiting a more powerful computer to solve; it is the methodological signature of an idea functioning regulatively rather than constitutively.

This is precisely the configuration Kant outlines for cosmological ideas,¹⁶ though scoped down to empirical totalities.¹⁷ Reason constantly strives to generate ideas of totality, such as the world as a complete whole (the cosmological system), but these can never be given in possible experience as determinate objects. When we treat them constitutively rather than regulatively, reifying the pointer into the thing itself, we produce *dialectical illusion*. The same structure recurs wherever a complex system exhibits genuine computational

¹⁵“The whole is the false.” Theodor W. Adorno, *Minima Moralia*, tr. E.F.N. Jephcott (London: Verso, 2005), 50.

¹⁶Kant, *Critique of Pure Reason*, A405–567/B432–595, 423–463.

¹⁷See A645–647/B673–675 (555–557) and A660–664/B688–692 (563–566), where Kant discusses how reason must operate regulatively in organising empirical knowledge into systematic unity precisely because it cannot determine this unity constitutively: ‘what reason quite uniquely prescribes and seeks to bring about [...] is the systematic in cognition, i.e., its interconnection based on one principle’ (A645/B673). For Kant’s account of the architectonic of pure reason as systematic unity under an idea, see A832–851/B860–879 (655–665), esp. A835/B863 (656). This description of the idea of *system* in general has underexplored consequences for the design and analysis of software.

irreducibility: the idea of the whole, whether a cosmological totality or a dynamical attractor, is what makes the manifold of particulars thinkable as a unity in the first place, yet it is never exhausted by any finite enumeration of those particulars.¹⁸ Dialectical illusion sets in only when this orienting idea is hypostatised into a constitutive object, mistaking the very function that renders the manifold intelligible for a determinate member of it.

Kant's distinction between *intellectus ectypus* and *intellectus archetypus* clarifies what is at stake when the regulative is mistaken for the constitutive. Human understanding is fundamentally *ectypal*: it can know only through representations, requiring the mediation of intuitions and concepts. By contrast, the *intellectus archetypus*, a regulative fiction, would be a form of cognition where thought and reality are one, where to think something would be for it to exist.¹⁹ As Kant writes, such an understanding would be one for which 'the representation of a whole contain[s] the ground of the possibility of its form',²⁰ where there is no gap between concept and object. Superintelligence, insofar as it designates anything coherent, gestures *toward* such an archetypal condition: intelligence without the constitutive limitations that define all possible forms of human cognition.



Today, there is of course great incentive²¹ to prematurely concretise the concept of AI or to offer unqualified prophecy. This futurological imperative

¹⁸This is not a peculiarity confined to the cosmological Antinomies but, according to a growing body of scholarship, a pattern constitutive of Kant's thought as a whole (if you will forgive me). See Lea Ypi, *The Architectonic of Reason: Purposiveness and Systematic Unity in Kant's Critique of Pure Reason* (Oxford: Oxford University Press, 2021), who argues that systematic unity functions as reason's organising ideal across the Critiques rather than in the Transcendental Dialectic alone; and Jennifer Mensch, *Kant's Organicism: Epigenesis and the Development of Critical Philosophy* (Chicago: University of Chicago Press, 2013), who shows the epigenetic model of the organism furnishing a structurally identical figure, in which the idea of the whole governs and unifies its parts without being reducible to their aggregate.

¹⁹This reverses the direction of travel Vico assigns to the history of language: for Vico, the earliest form of speech is a condition in which word and thing are flush, immediate and non-arbitrary, the "poetic" sign of the age of gods and heroes, from which humanity only gradually abstracts into the conventional, arbitrary signification of rational-civil discourse. The *intellectus archetypus* posits that same fusion of sign and referent not as a lost origin but as reason's own unreachable horizon, approached, if at all, from the far side of abstraction rather than before it.

²⁰Immanuel Kant, *Critique of the Power of Judgment*, ed. Paul Guyer, tr. Paul Guyer and Eric Matthews (Cambridge: Cambridge University Press, 2000), 277.

²¹See Ruchir Sharma, 'America is now one big bet on AI', *Financial Times*, 6 October 2025. Sharma argues that AI-related expenditure now accounts for approximately 40 per cent of US GDP growth in 2025, while AI companies have accounted for 80 per cent of gains in US stocks. [Update, October 2026: in May 2026 Goldman Sachs estimated that AI investment would account for roughly half of S&P 500 earnings growth this year.]

demands serious scrutiny. Lyotard describes the attempt to trim down the space of possible futures in terms of temporal engineering:

If one wants to control a process the best way of doing so is to subordinate the present to what is (still) called the future, since in these conditions the future will be completely predetermined and the present itself will cease opening onto an uncertain and contingent ‘afterwards’ [...] the present loses its privilege of being an ungraspable point from which, however, time should always distribute itself between the ‘not yet’ of the future and the no longer of the past.²²

This technique of subordinating the present to a predetermined future finds acute expression in contemporary AI discourse, where the regulative idea of superintelligence is converted as if by magic into a manageable technical problem. The proliferation of “AI safety” frameworks, alignment research programmes, and existential risk mitigation strategies participates in this temporal engineering exercise by treating an essentially indeterminate future as subject to present calculation and control. The economic pressures are obvious: venture capital demands roadmaps, policy requires threats it can name, and institutional research must justify its funding through claims of actionable foresight.

The AI ethicist cottage industry and the attendant alignment discourse exemplify this tendency. “Alignment” is a fundamentally incoherent concept because it immediately assumes human-like moral agency on the part of emerging technology, that is, free decisionality informed by held values or axiological systems. It presupposes we can specify a utility function for an agent whose very structure of preference may be incommensurable with ours, or whose operations may not admit of teleological interpretation at all. More subtly, it assumes that the future shape of advanced AI systems can be determined through present intervention, flattening Lyotard’s ‘uncertain and contingent “afterwards”’ into a tightly controlled space of technical optimisation.

Why would an emergent intelligence, whether of the “like us, but more so” or the “logical alien” type, necessarily replicate human moral reasoning?

²²Jean-François Lyotard, ‘Time Today’, *The Inhuman: Reflections on Time*, tr. Geoffrey Bennington and Rachel Bowlby (Stanford, California: Stanford University Press, 1991), 65.

Even were it to do so, could this not be an inscrutable move in an *abstract game* of the sort outlined above? If different in degree, its moves in such games may have scaled well past our comprehension. If different in kind, we may find ourselves locked out of its internally consistent system altogether. The alignment model is so reductive when faced with these intrinsic possibilities that it reads less as serious epistemology and more like a truculent control phantasy.



As much as Reason would love to extrapolate beyond its limited domain, we cannot see beyond what Ernst Jünger called the “wall of time” [*Zeitmauer*].²³ We should therefore distrust those peddling schlocky time-travel ethics and alignment gurus claiming special foresight into the behaviour of systems that they admit from the outset will exceed their own comprehension. We should be suspicious that AI doomerism and boosterism form a feedback loop that resembles capitalism’s longstanding ability to absorb its own direct critique into its apparatus of reproduction. In the words of Paul Valéry: ‘All prediction is conservative: it demands that we be as we are in whatever future it constructs.’²⁴ Often, the most adventurous flights of futurism mask a reactionary impulse to be sheltered from the radical contingency of what is to come.

Still, the strong deflationary position is equally untenable. To dismiss what has emerged from transformer architectures as a bag of statistical tricks is to wilfully ignore true discontinuities. The capacity to learn compressed representations of human linguistic production at this scale would have seemed fantastical mere decades ago. The frontier is already shifting: world models now generate interactive, physically grounded environments rather than passive video, operating over geometric and dynamical domains and not merely linguistic ones; a vendor-neutral protocol layer has standardised the business of calling external tools, so that models function as components in larger

²³ ‘The question then arises as to what kind of phenomena we can expect in the future. The answer is that *we can expect phenomena that are not yet known to us*. [...] Herodotus gazed back from the historical space he had just entered into the mythical one. He did so with trepidation. The same trepidation is called for today, where the future looms beyond the wall of time.’ Ernst Jünger, *An der Zeitmauer* (Stuttgart: Ernst Klett, 1981), 39.

²⁴ Paul Valéry, ‘Unpredictability’, in *The Outlook for Intelligence*, tr. Denise Folliot and Jackson Matthews (New York: Harper & Row, 1963), 69.

agentic systems rather than as terminal oracles;²⁵ coding assistants have begun to contribute materially to the production of their successors;²⁶ the pairing of language models with proof assistants has started to settle open mathematical questions; and, most consequentially for anyone who has to live in the world these systems are being released into, frontier models have become capable of discovering exploitable flaws in large codebases and chaining them together, a faculty they have begun to turn on cryptographic schemes themselves.²⁷ Whether applied to token sequences, spatial configurations, or multimodal data, the learning mechanisms that enable statistical compression may well prove crucial to whatever eventually constitutes AGI, even if current architectures are insufficient. We face something genuinely novel, the trajectory of which remains radically uncertain, but which, either way, is not reducible to eschatological fantasy or offhand dismissal.

Reasoning models in recent years represent a shift in training regime and in where compute is spent: rather than responding immediately, they perform extended “internal” deliberation, generating chains of thought before producing outputs. Whether this escapes the limitation regarding efficient generalisation is an open question,²⁸ since these models operate by passing acquired context forward through extended inference, essentially re-prompting themselves. They learn broader patterns of reasoning such as problem decomposition strategies, verification steps, and common failure modes. It remains to be seen whether the acquisition of these broad-brush patterns can be subject to the same critique as learned representations. Each step in this

²⁵The significant development is not that a model can call a function, or “make a tool call”, which it always could, but that a vendor-neutral protocol layer now exists for it to call through, the Model Context Protocol having been widely adopted for the purpose.

²⁶The laboratories have now begun describing their models as instrumental in their own construction, used to debug training runs and to write the bulk of the code they are built from; DeepMind’s AlphaEvolves found a method for multiplying two 4×4 complex matrices in 48 scalar multiplications, one fewer than Strassen’s long-standing 49. None of this is the runaway self-amplification of the folk mythology, and ‘recursive self-improvement’ claims a great deal more than the evidence will bear; but the systems are now participants in their own genealogy.

²⁷The capability is double-edged by construction, since the faculty that audits a codebase is the same as the faculty that attacks it. In cryptography it has so far broken what is proposed rather than what is deployed: in July 2026 an Anthropic model found an attack on HAWK, a candidate signature scheme then under evaluation by NIST, which its designers withdrew within days.

²⁸The evidence cuts both ways. Much of the deliberation appears ornamental: Daniel Scalena et al., ‘Beyond the Commitment Boundary’, arXiv:2606.13603 (2026), find that the answer typically stabilises well before the chain of thought ends, and that up to 55 per cent of it can be cut without loss. Yet paired with a proof assistant the same models produce results no corpus contained: George Tsoukalas et al., ‘Advancing Mathematics Research with AI-Driven Formal Proof Search’, arXiv:2605.22763 (2026), report nine open Erdős problems resolved autonomously.

mode of sequential processing inherits the constraints of that which preceded it, and it is this path dependence that lends the development its interest. It brings to mind Kant's division of all cognition into historical cognition, *cognitio ex datis*, and rational cognition, *cognitio ex principiis*.²⁹ The former is the realm of cognition via aggregation of detail, while the latter is that of employing the compact generality of the concept. Models in their present form clearly excel at historical cognition. An attempt to coordinate the two modes is underway.

Whatever becomes of deliberate attempts to compose such models into collectives of specialised agents,³⁰ multi-agent behaviour will arrive anyway, as a by-product rather than a design.³¹ The deeper language models embed themselves in economic activity, the denser the network of local-global exchanges through which they operate.³² In such a network 'the meaning of information is nothing other than the set of actions it triggers and controls'.³³ What emerges is not a singular superintelligence but collective intelligence at an unfamiliar scale: dialogic swarm cognition rather than machinic monotheism.

Beyond ectypal human intelligence, confined to cognition through the concept, and beyond the hyperectypal operations of language models, churning their way through representations-of-representations in the tombs,³⁴

²⁹'If I abstract from all content of cognition, objectively considered, then all cognition, considered subjectively, is either historical or rational. Historical cognition is *cognitio ex datis*, rational cognition, however, *cognitio ex principiis*. However a cognition may have been given originally, it is still historical for him who possesses it if he cognises it only to the degree and extent that it has been given to him from elsewhere, whether it has been given to him through immediate experience or told to him or even given to him through instruction (general cognitions).' Kant, *Critique of Pure Reason*, A836/B864, 657.

³⁰The record for deliberately engineered multi-agent systems is so far unimpressive; see Mert Cemri et al., 'Why Do Multi-Agent LLM Systems Fail?', arXiv:2503.13657 (2025); and, on reported gains that owe more to extra compute than to coordination, Dat Tran and Douwe Kiela, 'Single-Agent LLMs Outperform Multi-Agent Systems on Multi-Hop Reasoning Under Equal Thinking Token Budgets', arXiv:2604.02460 (2026). The point here concerns emergence rather than design.

³¹For a survey of the formalisms available for describing such coordination, see Thomas Murphy, 'A Survey of Agentic Coordination Formalisms', *abstract_machines*, 2 September 2026.

³²See Fernando Zalamea's category-theoretic model of local/global transits. William Lawvere suggested that dialectical or antinomical dyads could be modelled in this manner in his introduction to *Categories in Continuum Physics*. Naturally there is an abundance of applied category theory formalisms of AI systems under development, though the extent to which engineers in industry will engage with these seems limited.

³³Raymond Ruyer, *Cybernetics and the Origin of Information*, tr. Amélie Berger-Soraruff, Andrew Iliadis, Daniel W. Smith, and Ashley Woodward (London: Bloomsbury Publishing, 2023), 2.

³⁴'The realm of the dead is as extensive as the storage and transmission capabilities of a given culture.' Friedrich Kittler, *Gramophone, Film, Typewriter*, tr. Geoffrey Winthrop-Young and Michael Wutz (Stanford, California: Stanford University Press, 1999), 13.

stands the regulative idea of an absolute *intellectus archetypus*: cognition without mediation; thought without the gap between sign and referent. Thought without thoughts.³⁵ Transcendental critique demonstrates that an idea can be logically impossible and still serve a purpose. It may be that the concept of superintelligence ultimately admits of an ‘indispensably necessary regulative use, namely that of directing the understanding to a certain goal respecting which the lines of direction of all its rules converge at one point’.³⁶ The myth itself may coordinate individual thought, and, at the scale of the social system, productive forces, toward its own realisation, whether partial or complete.

Nonetheless—and here we must part ways with those who would cut the overwhelming prospect of artificial intelligence down to size by projecting their own finite intellect onto that which is supposed to exceed it—it must continue to be understood as ‘an idea (*focus imaginarius*)—i.e., a point from which the concepts of the understanding do not really proceed, since it lies entirely outside the bounds of possible experience’. The danger is that this is where ‘the deception [arises], as if these lines of direction were shot out from an object lying outside the field of possible empirical cognition (just as objects are seen behind the surface of a mirror)’.³⁷

The search for orienting ideas is not itself at fault, because, under the conditions of modernity, ‘the main aim in the modes of understanding [...] is *to hit upon something successfully*, to have a fate, since fatelessness, *δύσμορον*, is our weakness’.³⁸ The ideal engineer subordinates the means to the end, dragging the abstract into the concrete. But *ideas that become real*—self-fulfilling prophecies—bootstrap themselves from germinal possibilities. They do not issue from neurotically overdetermined fantasies that pinion that which they purport to predict. The error lies not in the bare use of the virtual but its illegitimate actualisation.

Despite the widespread epistemic confusion and moral panic brought

³⁵This formulation transposes Heidegger’s ontological difference between Being (*Sein*) and beings (*Seiendes*) onto thought: Thought in the singular names the ground, thoughts in the plural its determinations. See Martin Heidegger, *Being and Time*, tr. John Macquarrie and Edward Robinson (Oxford: Blackwell, 1962), 26 (H. 6). As Kyoto School philosopher Nishida Kitarō well understood, the act of judgement is deeply imbricated in the problem of the continuous and the discrete. See *Intuition and Reflection in Self-Consciousness*, tr. Valdo H. Viglielmo, Takeuchi Yoshinori and Joseph S. O’Leary (Nagoya: Chisokudo Press, 2020).

³⁶Kant, *Critique of Pure Reason*, 555.

³⁷Kant, *Critique of Pure Reason*, 555.

³⁸Friedrich Hölderlin, ‘Notes on the Antigone’, *Essays and Letters*, tr. Jeremy Adler and Charlie Louth (London: Penguin Books, 2009), 330.

about by the initial successes of the transformer architecture, signs of what may come are there for all to read. The internet is no longer simply a trap for the harvesting of data³⁹ but a laboratory for its recombination from a civilisation-scale data-exhaust.⁴⁰ We are passing through a metasystem transition, between regimes of signs.⁴¹ No matter the woolly predictions, the moral fantasias, the forms taken by our *foci imaginarii*, the bleating of dogmatic cheerleaders and naysayers... all of this noise, this frenzied standstill, will resolve into the nervous hush of anticipation before the raising of a curtain. Ideas alone can only draw us so far into the future. For beyond the wall of time, *something* will arise, dripping, from the sea of information.

...noch sind Lieder zu singen jenseits der Menschen.
...and there are still songs to be sung beyond mankind.⁴²



First published on 24 October 2026 at
thomas-murphy.com/post/essays/a-late-evening-in-the-mechanosphere/
Set in EB Garamond, Garamond-Math and Caudex.

³⁹ An anachronistic reinterpretation of the “platform-phase” in the evolution of the internet in terms of the preparation of training data becomes ever more tempting.

⁴⁰ Disappointingly, as many observers have noted, its current iteration has so far assumed the form of a *slopwave*.

⁴¹ ‘A crisis is the passage from one particular mode of functioning to another; a passage made perceptible by signs or symptoms. During a crisis, time seems to change its nature...’ Paul Valéry, ‘Remarks on Intelligence’, *The Outlook for Intelligence*, 72.

⁴² Paul Celan, ‘Fadensonnen’, *Fadensonnen* (Frankfurt am Main: Suhrkamp Verlag, 1968). The translation combines Pierre Joris, ‘Threadsuns’, in *Breathturn into Timestead: The Collected Later Poetry* (New York: Farrar, Straus and Giroux, 2014), and Michael Hamburger, in *Poems of Paul Celan* (New York: Persea Books, 1988).